

Design characteristics of food frequency questionnaires in relation to their validity.

Citation for published version (APA):

Molag, M. L., de Vries, J. H., Ocke, M. C., Dagnelie, P. C., van den Brandt, P. A., Jansen, M. C., van Staveren, W. A., & van 't Veer, P. (2007). Design characteristics of food frequency questionnaires in relation to their validity. *American Journal of Epidemiology*, 166(12), 1468-1478. <https://doi.org/10.1093/aje/kwm236>

Document status and date:

Published: 01/01/2007

DOI:

[10.1093/aje/kwm236](https://doi.org/10.1093/aje/kwm236)

Document Version:

Publisher's PDF, also known as Version of record

Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

www.umlib.nl/taverne-license

Take down policy

If you believe that this document breaches copyright please contact us at:

repository@maastrichtuniversity.nl

providing details and we will investigate your claim.

Practice of Epidemiology

Design Characteristics of Food Frequency Questionnaires in Relation to Their Validity

Marja L. Molag¹, Jeanne H. M. de Vries¹, Marga C. Ocké², Pieter C. Dagnelie³, Piet A. van den Brandt³, Margje C. J. F. Jansen⁴, Wija A. van Staveren¹, and Pieter van't Veer¹

¹ Department of Human Nutrition, Wageningen University, Wageningen, The Netherlands.

² National Institute of Public Health and the Environment, Bilthoven, The Netherlands.

³ Department of Epidemiology, Maastricht University, Maastricht, The Netherlands.

⁴ TNO Quality of Life, Zeist, The Netherlands.

Received for publication March 12, 2007; accepted for publication July 27, 2007.

The authors investigated the role of food frequency questionnaire (FFQ) design, including length, use of portion-size questions, and FFQ origin, in ranking subjects according to their nutrient intake. They also studied the ability of the FFQ to detect differences in energy intake between subgroups and to assess energy and protein intake. In a meta-analysis of 40 validation studies, FFQs with longer food lists (200 items) were better than shorter FFQs at ranking subjects for most nutrients; results were statistically significant for protein, energy-adjusted total fat, and energy-adjusted vitamin C. The authors found that FFQs that included standard portions had higher correlation coefficients for energy-adjusted vitamin C (0.80 vs. 0.60, $p < 0.0001$) and protein (0.69 vs. 0.61, $p = 0.03$) than FFQs with portion-size questions. However, it remained difficult from this review to analyze the effects of using portion-size questions. FFQs slightly underestimated gender differences in energy intake, although level of energy intake was underreported by 23% and level of protein intake by 17%. The authors concluded that FFQs with more items are better able to rank people according to their intake and that they are able to distinguish between subpopulations, even though they underestimated the magnitude of these differences.

diet; methods; nutrition assessment; questionnaires; review; validation studies

Abbreviations: EPIC, European Prospective Investigation into Cancer and Nutrition; FFQ, food frequency questionnaire; SE, standard error.

Food frequency questionnaires (FFQs) are widely used to assess the dietary intake of large populations. The popularity of FFQs stems from their ease of administration, ability to assess dietary intake over an extended period of time, and low costs (1). They are therefore often used in epidemiologic studies to investigate the relation between diet and disease. For some purposes, information about level of intake is very important, for example, to set recommendations for nutrient intake. For most epidemiologic studies, FFQs must be able to classify individuals correctly according to

their dietary intake. However, Bingham and others have argued that, probably because of misclassification, FFQs are not always able to detect weak associations (2, 3). Because of this debate and their established role in epidemiologic research, FFQs need to be further characterized and subsequently improved.

FFQs differ in the way they are developed and show large variations in design characteristics, such as number of items or inclusion of portion-size questions. Such variations could affect reported intakes (4–6).

Correspondence to Dr. J. H. M. de Vries, Department of Human Nutrition, Wageningen University, P.O. Box 8129, 6700 EV Wageningen, The Netherlands (e-mail: Jeanne.deVries@wur.nl).

In this study, we first aimed to provide an overview of FFQ validation studies and the validity of FFQs in classifying subjects relative to their intake in relation to number of items in the FFQ, use of portion-size questions, origin of the FFQ, and administration mode. A second aim was to provide an overview of the validity of FFQs in assessing absolute energy and protein intakes as determined in studies using recovery biomarkers and to establish whether FFQs can detect known differences in energy intake between men and women.

MATERIALS AND METHODS

Search strategy and study selection

We searched MEDLINE (National Library of Medicine, Bethesda, Maryland) for validation studies of FFQs that assessed respondents' habitual dietary intake and were published between 1980 and December 2006. An FFQ was defined as a questionnaire with a food list and a frequency-response section where subjects report how often each food item is consumed (7, p. 75). Search terms used were the Medical Subject Headings (MeSH) "nutrition assessment," "questionnaires," "evaluation studies" OR "reproducibility of results" AND keywords "food frequency questionnaire" OR "FFQ" AND "validity" OR "validation" OR the publication type "validation studies."

Studies that met all of the following inclusion criteria were included in the review:

- 1) Describing FFQs developed for epidemiologic purposes
- 2) Addressing habitual diet by reporting at least energy intake but preferably also intake of other nutrients including total fat, carbohydrates, protein, alcohol, calcium, and vitamin C
- 3) Studying adult populations in the age range 18–82 years with a westernized diet but not FFQs validated exclusively among those older than age 60 years
- 4) Validating FFQs with one of the following reference methods: 24-hour dietary recalls, food records, diet history interview, or recovery biomarkers

Studies that assessed only specific nutrients or food groups such as fruit and vegetable consumption were excluded from this review.

Data extraction and classification

The following design characteristics of the FFQ were extracted: number of items in the FFQ (ranging from 44 to 350), use of portion-size questions versus predefined standard portion sizes, and "origin" of the FFQ, for example, the Willett type (8), the Block type (9), the European Prospective Investigation into Cancer and Nutrition (EPIC) type (10), and "other FFQs." We extracted the following validation study characteristics: study size (*N*), gender (males, females), FFQ administration mode (interview or self-administration), reference method (food record, 24-hour dietary recall, or diet history interview), and total number of days over which the reference method was applied—

categorized as a short period of 1–7 days, a medium period of 8–14 days, and a long period of 15 days or more. We used 8–14 days as the reference category in the meta-regression analyses because the 24-hour dietary recall method was not applied for 15 days or more.

We analyzed nutrients that covered different aspects of a habitual diet. We included energy and all energy-providing macronutrients (i.e., protein, carbohydrates, fat, and alcohol). We added vitamin C and dietary fiber to represent fruit and vegetables and some other food components, for example, calcium, to represent other specific foods.

Statistical analysis

Ranking. We extracted crude and energy-adjusted correlation coefficients and, if available, gender-specific and/or gender-adjusted correlation coefficients between FFQs and reference methods to compare studies with respect to ranking of subjects. Correlation coefficients were first converted into a standard normal metric by using Fisher's *r*-to-*Z* transformation, expressed in equation 1 (11), in which r_i is the correlation coefficient from study *i*.

$$Zr_i = \frac{1}{2} \text{Log}_e \left(\frac{1+r_i}{1-r_i} \right). \quad (1)$$

The transformed effect sizes were then used to calculate an initial pooled average, mean Zr_i , in which each correlation coefficient was weighted by the variance, $1/(n_i - 3)$. Then, to identify confounders to be used for adjustments in meta-regression analyses, we stratified these correlation coefficients by sex, type of reference method, category of number of days that this method was applied, and administration method. The weighted averages, mean Zr , were also used to perform the Cochran *Q* test for heterogeneity between studies for the assessed nutrients, expressed in equation 2 (12).

$$Q = \sum_{i=1}^k (n_i - 3)(Zr_i - \text{mean } Zr). \quad (2)$$

This test had a *p* value of <0.0001 for all nutrients, which indicated that the correlation coefficients from the validation studies were heterogeneous.

Following the heterogeneity test, random-effects meta-regression was conducted to explain this heterogeneity by FFQ design characteristics using the restricted maximum likelihood approach, as per Thompson and Sharp (13). For each nutrient, number of items as a continuous variable, use of portion-size questions or standard portions, FFQ origin—Willett, Block, EPIC, or other—and several potential confounders (gender, reference method, and number of days over which this method was applied) were regressed on the transformed correlation coefficients Zr_i . Weights were assigned based on the variance ($1/(n_i - 3)$). Results of the meta-regression are presented as predicted values of *Z*, retransformed to *r*, using a model that included an intercept, a reference period of 8–14 days, an average value of 0.5 for the indicator variable for sex, and similarly so for the

TABLE 1. Overview of the 40 studies included in the review by first author, year of publication, and “origin” of the FFQ*

First author, year (reference no.)	Origin†	No. of subjects	No. of FFQ items	FFQ administration method	Use of portion-size questions	FFQ reference period	Reference method	Repetitions reference method
Andersen, 1999 (26)	Other	125 men	180	Self	Yes	Unknown	FR*	14 d* in 5 w*
Andersen, 2003 (53)	Other	17 women	180	Self	Yes	1 y*	DLW*	1 (10 d)
Bingham, 1997 (27)	EPIC*	156 women	130	Self	No	Unknown	FR	4 × 4 d in 1 y
Block, 1990 (28)	Block	102 women	60	Self	Yes	Unknown	FR	2 × 7 d
Block, 1990 (9)‡	Block	102 women	94	Self	Yes	6 m*	FR	3 × 4 d in 1 y
Block, 1992 (29)	Block	85 men and women	98	Interview	Yes	1 y	FR	4 × 3 d in 1 y
Bohlscheid-Thomas, 1997 (56)	EPIC	49 men, 55 women	104	Self	Yes	1 y	24-HR*	12× in 1 y
Boucher, 2006 (17)	Block	166 women	126	Self	Yes	Unknown	24-HR	2×
Brunner, 2001 (30)	Other	457 men, 403 women	127	Self	No	1 y	FR	7 d
Callmer, 1993 (31)	Other	57 men, 50 women	250	Self	Yes	1 y	FR	6 × 3 d in 1 y
Engle, 1990 (32)	Other	16 men, 34 women	120	Self	Yes	3 m	FR	7 d
Feunekes, 1993 (50)	Other	95 men, 96 women	104	Interview	Yes	1 m	DH*	1 × 1 m
Fidanza, 1995 (33)	Other	11 men, 35 women	93	Self	Yes	1 y	FR	7 d
Flagg, 2000 (18)	Block	216 men, 223 women	114	Self	Yes	1 y	24-HR	4× in 1 y
Goldbohm, 1994 (34)	Other	59 men, 50 women	150	Self	Yes	1 y	FR	3 × 3 d in 1 y
Hartwell, 2001 (35)	Other	16 men, 9 women	162	Self	No	1 y	FR	2 × 4 d
Jain, 2003 (19)	Other	151 men, 159 women	166	Self	Yes	1 y	24-HR	3× in 1 y
Johansson, 2002 (20)	EPIC	96 men, 99 women	84	Interview	Yes	Unknown	24-HR	10× in 1 y
Katsouyanni, 1997 (21)	EPIC	42 men, 38 women	190	Self	Yes	1 y	24-HR	12× in 1 y
Kroke, 1999 (51)	EPIC	75 men, 59 women	146	Self	Yes	1 y	DLWS, 24-HR	14 d, 12× in 1 y
Larkin, 1989 (36)	Other	228 men and women	116	Interview	Yes	1 y	FR	16 d in 1 y
Longnecker, 1993 (37)	Willett	64 men, 74 women	116	Self	No	1 y	FR	3 × 2 d or 2 × 2 d
Mannisto, 1996 (38)	Other	152 women	110	Self	No	1 y	FR	2 × 7 d in 3 m

Table continues

reference method. All data were analyzed by using STATA 8 software (Stata Corporation, College Station, Texas).

Validity of absolute intake. To assess energy intake, validity was defined as the difference between the mean levels of energy intake assessed by the FFQ minus the mean levels of energy expenditure determined by the doubly labeled water method. For protein intake, it was defined as the difference between protein intake assessed by the FFQ and protein intake estimated from 24-hour urinary nitrogen excretion. If only nitrogen excretion was reported in the paper, we estimated protein intake by assuming that urinary nitrogen was excreted as a constant proportion of 80 percent of total nitrogen intake (14), and 16 percent of protein is ni-

trogen (15). Thus, protein intake was estimated from the following formula:

$$\text{protein} = 6.25 \times (\text{urinary } N / 0.80). \quad (3)$$

Validity of gender differences in energy intake. To evaluate the extent to which gender differences in energy intake could be detected by FFQs, we extracted gender-specific mean energy intake, including standard deviations (if not available, we assumed it was 3 MJ because this mean of standard deviations was reported in 31 other included studies) and *N*, or the standard error (SE) for mean energy

TABLE 1. Continued

First author, year (reference no.)	Origin†	No. of subjects	No. of FFQ items	FFQ administration method	Use of portion-size questions	FFQ reference period	Reference method	Repetitions reference method
Martin-Moreno, 1993 (39)	Other	147 women	118	Self	Yes	1 y	FR	4 × 4 d in 1 y
McKeown, 2001 (40)	EPIC	58 men, 88 women	130	Self	No	1 y	FR	2 × 7 d
Munger, 1992 (41)	Willett	44 women	126	Self	No	Unknown	FR	3 d
Ocke, 1997 (22)	EPIC	63 men, 58 women	178	Self	Yes	1 y	24-HR	12× in 1 y
Patterson, 1999 (42)	Other	113 women	122	Self	Yes	3 m	FR	4 × 4 d
Pietinen, 1988 (43)	Other	190 men	203	Self	Yes	1 y	FR	12 × 2 d in 6 m
Pietinen, 1988 (44)	Other	297	44	Self	No	1 y	FR	12 × 2 d in 6 m
Pisani, 1997 (23)	EPIC	47 men, 150 women	47	Self	Yes	Unknown	24-HR	8–14× in 1 y
Riboli, 1997 (45)	EPIC	57 men, 50 women	350	Self	Yes	1 y	FR	6 × 3 d in 1 y
Rimm, 1992 (46)	Willett	127 men and women	131	Self	No	1 y	FR	2 × 7 d in 6 m
Schroder, 2001 (47)	Other	44 men and women	157	Self	Yes	Unknown	FR	3 d
Subar, 2001 (24)								
Willett	Willett	254 men, 293 women	126	Self	No	1 y	24-HR	4× in 1 y
DHQ*	Other	501 men, 560 women	124	Self	Yes	1 y	24-HR	4× in 1 y
Block	Block	247 men, 267 women	106	Self	Yes	1 y	24-HR	4× in 1 y
Subar, 2003 (52)	Other (DHQ)	261 men, 223 women	124	Self	Yes	1 y	DLW, 24-HR	14 d, 2× in 3 m
Tjonneland, 1991 (48)	EPIC	59 men, 85 women	92	Self	Yes	1 y	FR	2 × 7 d
van Liere, 1997 (25)	EPIC	123 women	238	Self	Yes	Unknown	24-HR	12× in 1 y
Willett, 1985 (8)	Willett	173 women	61	Self	No	1 y	FR	4 × 7 d in 1 y
Willett, 1988 (49)	Willett	150 women	116	Self	No	1 y	FR	4 × 7 d, 3–4 y before

* FFQ, food frequency questionnaire; FR, food record; d, days; w, weeks; y, year; DLW, doubly labeled water; EPIC, European Prospective Investigation into Cancer and Nutrition; m, months; 24-HR, 24-hour dietary recall method; DH, diet history; DHQ, dietary history questionnaire.

† Refers to FFQs developed from, for example, the Willett type (8), the Block type (9), the EPIC type (10), and “other FFQs.”

‡ Only women from the usual diet group were included in our analyses.

§ DLW was used in only a subgroup of 28 subjects.

intake. We subtracted the mean level of energy intake of the women estimated by each FFQ (FFQ_{women}) from that of the mean level of the men (FFQ_{men}), and we did the same for the reference method (Ref_{women} and Ref_{men}).

Gender difference FFQ:

$$FFQ_{\text{men}} - FFQ_{\text{women}} \pm \sqrt{SE_{FFQ_{\text{men}}}^2 + SE_{FFQ_{\text{women}}}^2} \quad (4)$$

Gender difference reference method:

$$Ref_{\text{men}} - Ref_{\text{women}} \pm \sqrt{SE_{Ref_{\text{men}}}^2 + SE_{Ref_{\text{women}}}^2} \quad (5)$$

We tested whether there was a difference between both results by using an independent *t* test.

RESULTS

Description of studies

The search procedure resulted in 40 papers (table 1) describing 42 FFQs that matched the inclusion criteria. The majority of FFQs were validated against 24-hour dietary recalls (16–25) or food records (8, 9, 26–49). One FFQ was validated against a diet history method (50), two FFQs against 24-hour dietary recalls and doubly labeled water (51, 52), and one FFQ against doubly labeled water only (53). Six FFQs (8, 24, 37, 41, 46, 49) were developed from the Willett FFQ (8). For this FFQ, an extensive food list was shortened by removing infrequently eaten items and including items contributing most to between-person variance

using data from Nurses' Health Study participants. Willett FFQs included on average 113 (range, 61–131) items and asked respondents to report their frequency of consumption of a given reference portion size in a table format. Another six FFQs (9, 17, 18, 24, 28, 29) were developed from the Block FFQ (9). This FFQ was developed by using food items that contributed over 90 percent of the total population intake of energy and several nutrients in the Second National Health and Nutrition Examination Survey database (54). These Block FFQs consisted of an average of 100 items (range, 60–126), and all asked portion-size questions.

Within the EPIC, project country-specific FFQs were developed including items that cumulatively contributed most to between-person variance (21, 23, 40, 48, 51, 55, 56) or to total nutrient intake (22); of three FFQs, the method of development was not described (20, 25, 45). For the EPIC FFQs, we found 11 validation studies performed in nine countries (16, 20–23, 25, 27, 40, 45, 48, 51, 56); two were conducted in the United Kingdom (27, 40) and two in Germany (16, 51, 56). We analyzed them as separate studies. The FFQs validated in the EPIC studies consisted of an average of 154 items (range, 47–350). Nine of these FFQs included portion-size questions, and two assigned standard portion sizes. Although FFQ design between EPIC FFQs varied a lot, the design of their validation studies (57) was carefully standardized, except for the United Kingdom, Denmark, and Sweden because they joined the EPIC project at a later stage (27, 45, 48, 57). Three other EPIC validation studies did not match the inclusion criteria (58–60).

The “other FFQs” were also developed by including items contributing most to between-person variance or to total population intake. We included 19 FFQs as “other FFQs” (19, 24, 26, 30–36, 38, 39, 42–44, 47, 50, 52, 53). They consisted of an average of 139 items (range, 44–250); 15 of them included portion-size questions, and four assigned standard portion sizes.

Between-study differences in ranking subjects

For all nutrients, pooled correlation coefficients between FFQ and reference methods ranged from 0.45 for energy and protein to 0.74 for alcohol (table 2), and energy-adjusted correlation coefficients were 0.02–0.08 higher for most nutrients, except for vitamin C (0.05 lower). There were differences between studies due to gender and the reference method used, although they were not statistically significant (table 2). As expected, for most nutrients, correlation coefficients were significantly higher when the reference method was used for 8–14 days than for 1–7 days (table 2). Correlation coefficients did not increase further when the reference method was used for 15 days or more. For all nutrients, we also looked at the number of consecutive days on which a food record was kept and found that correlation coefficients were lower when the reference method consisted of food records kept for more than 5 days consecutively. After energy adjustment, these differences became less pronounced or even reversed.

We observed no statistically significant differences in correlation coefficients between the interviewer- (19, 20, 29,

36) and self-administered FFQs regarding the nutrients considered (table 2).

Heterogeneity by FFQ design characteristics

In the meta-regression analyses, we observed that FFQs with longer food lists (200 items) had 0.01–0.17 higher correlation coefficients than FFQs with shorter food lists (100 items) for most nutrients (table 3). Correlation coefficients were even higher for longer food lists for crude protein (0.56 for 200 items vs. 0.46 for 100 items, $p = 0.002$), energy-adjusted protein (0.68 vs. 0.51, $p < 0.001$), energy-adjusted total fat (0.68 vs. 0.59, $p = 0.02$), and energy-adjusted vitamin C (0.68 vs. 0.51, $p = 0.001$; table 3). The diet history method was used in only one study and therefore was excluded from main analyses.

FFQs with portion-size questions had much higher correlation coefficients for energy-adjusted alcohol than FFQs with predefined standard portions (0.76 vs. 0.61), although they were not statistically significant ($p = 0.23$). On the contrary, FFQs with portion-size questions had significantly lower correlation coefficients for energy-adjusted protein (0.61 vs. 0.69, $p = 0.03$) and for energy-adjusted vitamin C (0.60 vs. 0.80, $p < 0.001$) than FFQs with predefined standard portions. For other nutrients, correlation coefficients were 0.08 lower to 0.08 higher for FFQs with portion-size questions compared with standard portions.

Regarding origin of the FFQ, we observed that correlation coefficients for most crude macronutrients were higher for Block and EPIC FFQs than for Willett and “other” FFQs. For calcium and vitamin C, Willett FFQs performed better, and, after energy adjustment, other correlation coefficients improved for Willett FFQs.

Validity in absolute intake

Absolute energy intake estimated by FFQs was validated against energy expenditure estimated with the doubly labeled water method (51–53). Two FFQs in small European studies underestimated energy intake by 11 percent and 19 percent (51, 53). In one large study conducted in the United States, energy intake was underestimated by 34 percent for men and 36 percent for women (52), and protein intake was underestimated by 32 percent for men and 29 percent for women. Five EPIC FFQs were also validated against level of protein intake estimated from urinary nitrogen excretion (22, 27, 40, 45, 51, 52) (table 4). In these studies, estimation of protein intake varied from an underestimate of 23 percent to one study that overestimated protein intake by 18 percent for men and 25 percent for women (61). In the latter study, the longest FFQ with 350 items was applied.

Gender differences in energy intake

Because the goal of FFQs is to distinguish subpopulations that differ with respect to nutrient intake, we tested whether FFQs are able to detect a “known” difference in energy intake between men and women. On average, the gender difference in energy intake was smaller according to FFQs (2.09 MJ for men minus women, 95 percent confidence interval: 1.62, 2.56) than according to the reference methods

TABLE 2. Pooled correlation coefficients† for energy and nutrients between FFQs and reference methods stratified by characteristics of the validation studies

	No. of FFQs‡	Energy intake	Total fat	Protein	Carbohydrate	Alcohol	Calcium	Vitamin C	Dietary fiber
No. of FFQs‡	40	40	35	30	32	21	25	27	32
Crude <i>r</i> (range)	40	0.45 (0.16–0.77)	0.53 (0.18–0.88)	0.45 (0.14–0.70)	0.53 (0.25–0.77)	0.74 (0.29–0.90)	0.55 (0.20–0.75)	0.58 (0.16–0.82)	0.46 (0.25–0.74)
Validation study characteristics									
Reference method									
Diet history	1	0.77	NA§	NA	NA	NA	NA	NA	NA
24-HR§	13	0.42	0.49	0.46	0.56	0.69	0.55	0.55	0.48
FR§	26	0.53	0.56	0.47	0.51	0.78	0.56	0.59	0.44
FR <6 consecutive days	16	0.56	0.59	0.47	0.53	0.78	0.59	0.63	0.49
FR ≥6 consecutive days	10	0.42	0.51	0.46	0.50	0.78	0.53	0.55	0.36
No. of days that reference method was applied									
1–7	13	0.37	0.59	0.33	0.45	0.69	0.54	0.64	0.41
8–14	17	0.48*	0.54	0.51*	0.58*	0.76	0.57	0.56	0.46
≥15	10	0.46*	0.44	0.49*	0.55	0.80	0.59	0.52	0.56
Population characteristics									
Gender									
Men	19	0.44	0.52	0.49	0.57	0.66	0.56	0.53	0.44
Women	26	0.38	0.47	0.45	0.52	0.75	0.57	0.60	0.52
Both, unadjusted	10	0.48	0.46	0.45	0.48	0.82	0.45	0.46	0.42
Administration method									
Interviewer	4	0.50¶	0.50¶	0.44¶	0.58¶	0.54¶	0.51¶	0.51¶	0.54¶
Self	35	0.43	0.54	0.45	0.53	0.75	0.56	0.59	0.46
No. of FFQs‡	22		19	17	18	13	14	17	16
Energy-adjusted <i>r</i> (range)			0.56 (0.27–0.82)	0.49 (0.16–0.76)	0.61 (0.36–0.75)	0.76 (0.32–0.90)	0.61 (0.39–0.71)	0.53 (0.19–0.79)	0.54 (0.24–0.74)
Validation study characteristics									
Reference method									
24-HR	7		0.54	0.47	0.57	0.74	0.65¶	0.55	0.48
FR	15		0.57	0.51	0.57	0.78	0.59	0.52	0.59
FR <6 days consecutive	7		0.57	0.46	0.57	0.80	0.63	0.42	0.59
FR ≥6 consecutive days	8		0.57	0.53	0.58	0.76	0.58	0.54	0.51
No. of days that reference method was applied									
1–7	6		0.61¶	0.22¶	0.50	0.43¶	0.59¶	0.54¶	0.39¶
8–14	10		0.55	0.55*	0.62	0.81*	0.64*	0.54	0.55
≥15	6		0.55	0.45	0.53	0.84*,¶	0.62	0.49	0.61
Population characteristics									
Gender									
Men	8		0.51	0.50	0.60	0.72	0.64	0.49	0.39
Women	13		0.58	0.45	0.55	0.78	0.61	0.59	0.54
Both, unadjusted	5		0.62¶	0.59¶	0.60¶	0.75¶	0.57¶	0.35¶	0.61¶

* $p < 0.05$ for this characteristic by meta-regression analyses using a model that contained only this characteristic.

† Results pooled as Z-transformed values, weighed by variance, retransformed to correlation coefficients between food frequency questionnaires (FFQs) and reference methods.

‡ Maximum number of FFQs in this category; not all FFQs included all nutrients.

§ NA, not available; 24-HR, 24 hour dietary recall method; FR, food record.

¶ Based on <5 FFQs.

TABLE 3. Correlation coefficients† calculated from predicted values in a meta-regression model for energy and nutrients between FFQs and reference methods by characteristics of the FFQ

	No. of FFQs‡	Energy intake	Total fat	Protein	Carbohydrate	Alcohol	Calcium	Vitamin C	Dietary fiber
<i>Crude correlation coefficients</i>									
No. of FFQs‡	39	39	35	30	32	21	25	27	32
Design characteristics									
FFQ size§									
100 items		0.52	0.51	0.46	0.55	0.80	0.58	0.56	0.52
150 items		0.52	0.54	0.51*	0.57	0.80	0.60	0.60	0.53
200 items		0.53	0.56	0.56*	0.58	0.80	0.63	0.64	0.53
Portions¶									
Standard portions	11	0.46	0.52	0.55	0.56	0.72	0.67	0.68	0.45
Portion-size questions	28	0.52	0.54	0.51	0.57	0.80	0.60	0.60	0.53
Origin#									
Willett	6	0.51	0.42	0.46	0.49	0.36††	0.64††	0.67††	0.47
Block	6	0.58	0.53	0.50††	0.56	0.42††	0.56††	0.63††	0.58
EPIC	11	0.48	0.53	0.49	0.55	0.82*	0.52††	0.53	0.44
Other	16	0.54	0.48	0.44	0.52	0.83*	0.53	0.48*	0.55
<i>Energy-adjusted correlations</i>									
No. of FFQs‡	22		19	17	18	13	14	17	16
Design characteristics									
FFQ size§									
100 items			0.59	0.51	0.59	0.68	0.68	0.51	0.54
150 items			0.64*	0.61*	0.63	0.76	0.69	0.60*	0.57
200 items			0.68*,††	0.68*,††	0.66††	0.81††	0.70††	0.68*,††	0.59††
Portions¶									
Standard portions	9		0.66	0.69	0.63	0.61	0.68	0.80	0.50
Portion-size questions	12		0.64	0.61*	0.68	0.76	0.69	0.60*	0.57
Origin#									
Willett	5		0.52	0.47	0.61	0.36	0.68	0.69	0.56
Block	2		0.62††	NA††,‡‡	0.63††	0.52††	NA††	NA††	0.62*,††
Epic	5		0.64	0.61	0.59	0.82*	0.63	0.50	0.55
Other	10		0.42	0.56	0.58	0.85*	0.66	0.52*	0.61

* $p < 0.05$ in meta-regression analyses.

† Results of the meta-regression are presented as predicted values of Z , retransformed to correlation coefficients between food frequency questionnaires (FFQs) and reference methods, using a model that included the intercept and a reference period of 8–14 days, and at an average value of 0.5 for the indicator variable for gender and also at an average value of 0.5 for the reference method.

‡ Maximum number of FFQs in this category; not all FFQs included all nutrients.

§ Including the use of portion-size questions in the meta-regression model; 100 items was used as the reference.

¶ Including 150 items on the FFQ in the meta-regression model; standard portions were used as the reference.

Refers to FFQs developed from, for example, the Willett type (8), the Block type (9), the EPIC type (10), and "other FFQs"; the Willett FFQ was used as the reference.

†† Based on <5 FFQs.

‡‡ NA, not available.

(2.62 MJ, 95 percent confidence interval: 2.24, 3.00). Thus, on average, FFQs underestimated the gender differences in energy intake compared with the reference methods (figure 1) by -0.53 MJ (95 percent confidence interval: -1.13 , 0.07 , $p = 0.09$ using an independent t test). Exceptions were the two longest FFQs (31, 45), consisting of 250 and 350 items, respectively; they overestimated the difference in energy intake between men and women. The three FFQs that did

not include portion-size questions (24, 30, 40) found on average a smaller gender difference than FFQs that asked portion-size questions (18–24, 31, 45, 48).

DISCUSSION

This quantitative review of studies validating FFQs shows that the number of items in the food list is the major

TABLE 4. Mean energy intake reported by the FFQs* compared with mean energy expenditure determined by doubly labeled water, and mean protein intake reported by the FFQs compared with protein intake calculated from nitrogen excreted in urine

First author, year (reference no.)	No. of subjects	Mean intake reported by the FFQ	Reference method	Repetitions reference method	Underestimation in FFQ (%)
Doubly labeled water		Energy intake (MJ/day)	Energy expenditure (MJ/day)		
Andersen, 2003 (53)	17 women	8.28	9.23	1 (10 d*)	11
Kroke, 1999 (51)	28 men and women	9.05	11.23	1 (14 d)	19
Subar, 2003 (52)	245 men, 206 women	7.90 for men, 6.11 for women	11.92 for men, 9.53 for women	1 (11–14 d)	34 for men, 36 for women
Protein		Protein intake (g/day)	Protein excretion (g/day)		
Kroke, 1999 (51)	75 men, 59 women	75	97	4× in 1 y*	23
Ocke, 1997 (22)	46 men, 43 women	92 for men, 71 for women	99 for men, 81 for women	4× in 1 y	7 for men, 12 for women
McKeown, 2001 (40)	57 men, 77 women	89 for men, 78 for women	100 for men, 77 for women	6× in 9 m*	18 for men, 12 for women
Riboli, 1997 (45)	29 men, 24 women	95 for men, 76 for women	78 for men, 62 for women	8× in 1 y	–18 for men, 25 for women
Bingham, 1997 (27)	156 women	82	77	4 × 2 d in 1 y	–6
Subar, 2003 (52)	202 men, 150 women	71 for men, 55 for women	104 for men, 77 for women	2× in 14 d	32 for men, 29 for women

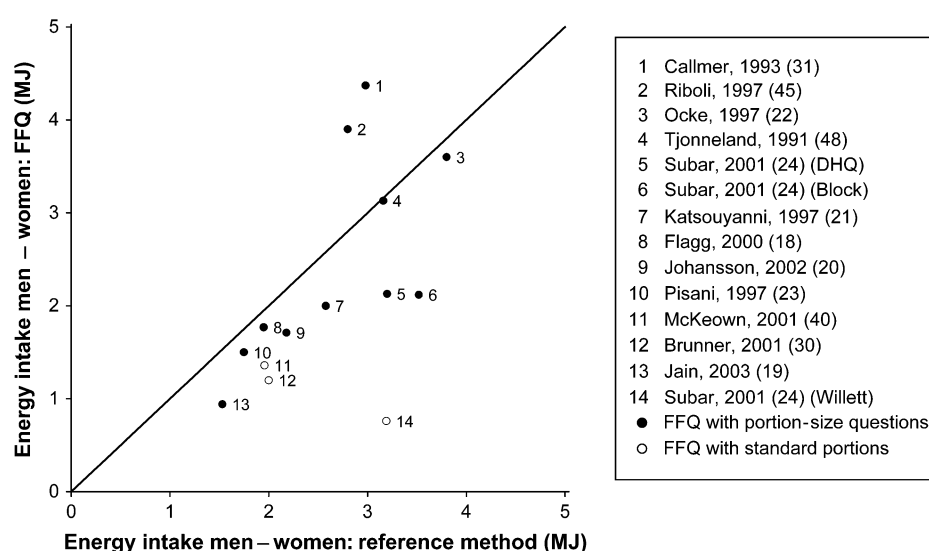
* FFQs, food frequency questionnaires; d, days; y, year; m, months.

determinant in ranking subjects with respect to their intake (for 100 extra items, correlation coefficients increased by 0.01–0.13). In general, portion-size questions and FFQ origin influenced ranking of subjects only slightly.

An important point of discussion is comparability of the studies included in this review. We aimed to address differences in characteristics of FFQ design and not of study design. We increased comparability by restricting our analyses to FFQs developed to cover the complete diet and validated among adults with Western food habits, and we adjusted for

potential confounders such as the reference method and the total number of days this method was used. There were differences in the design of the validation study: correlation coefficients were lower when food records were used for more than 5 days consecutively. However, this finding did not influence our results because there were also numerous studies included with another design.

In this study, we were not able to adjust for differences in energy needs or underreporting related to body weight and physical activity because these data were available for only

**FIGURE 1.** Mean energy intake of men minus mean energy intake of women according to the food frequency questionnaire (FFQ) and reference method including the 24-hour dietary recall method or food records. The boxed key lists the first author's name, year of publication, and reference number of the study corresponding to the numbered circles in the figure. DHQ, diet history questionnaire.

four studies. To account for this limitation, we evaluated both crude and energy-adjusted correlation coefficients because energy adjustment leads to a focus on the relative composition of the dietary pattern (62), and it has been suggested that it reduces correlated errors between the FFQ and reference methods (63).

In addition, variation in FFQ design was limited: the FFQs varied in the number of items and the use of portion-size questions, but differences in the reference period and the administration method were limited, prohibiting conclusions regarding the latter. Finally, we accounted for unknown between-study differences originating from study design and population by using a random-effects model in the meta-regression.

Our analyses showed that FFQs with a longer food list (200 items) were better at ranking people than FFQs with a shorter food list (100 items). These findings were clearest for protein and total fat, which are calculated from many different food sources. In the development phase of an FFQ, similar items are grouped together into items whose composition can be heterogeneous; an example is 20 different meat items combined into two items on the FFQ. Sometimes, items that contribute not much to total intake are omitted although they were important in explaining between-person variance. In summarizing, our results regarding the number of items should be used as an argument not to reduce the length of the food list too much when developing FFQs to rank persons according to nutrient intake.

Results of the meta-regression analyses showed that inclusion of portion sizes did not consistently affect the ranking of different nutrients. Ranking was worse for protein and vitamin C determined by FFQs that used portion-size questions instead of standard portions, and ranking improved for alcohol when FFQs used portion-size questions. An explanation for this unexpected finding might be that, for some foods such as vegetables, it is difficult to indicate how much was eaten, especially when they are part of mixed dishes (64). It might be easier to quantify the number and amount of alcoholic drinks; alcohol intake is ranked relatively well compared with other nutrients (65).

An important disadvantage of using standard portions is that interindividual variance decreases (66, 67). However, two validation studies in Denmark and the Netherlands found only small differences when analyzing FFQs using information from portion-size questions compared with analyzing the same FFQs using standard portions (68, 69). These small differences may reflect that quantification of portion sizes is of minor importance compared with frequency, that the relevant individual portion sizes were not estimated correctly (69), or that portion sizes listed do not correspond well with portions actually consumed. For example, actual portion sizes (e.g., super size) are probably much larger than standard portions used by US Department of Agriculture and the Food and Drug Administration (70, 71). Portion sizes were also estimated in different ways in the FFQs analyzed by including photographs, descriptions, and household measures such as spoons. Thus, it must be taken into account that portion-size questions do not always improve the performance of FFQs or that methods to estimate portion sizes should be improved.

The novelty of this review compared with previous reviews of FFQs (7, 67, 68, 72) is that we specifically analyzed the association between design characteristics of FFQs and their validity. Three other studies specifically compared the validity of Block- and Willett-type FFQs (24, 73, 74). A limitation of our review is that we could not disentangle the effects of type of questionnaire—Block or Willett FFQs—from the effects of use of portion-size questions and number of items. We did not have enough power to do so because only six Block and six Willett FFQs were included in the models. In general, we found that the Block FFQ performed better than the Willett FFQ, but, after energy adjustment, results regarding the different types were more comparable. This finding was also observed previously (24).

Apart from ranking subjects for etiologic studies, FFQs are sometimes used to assess absolute level of intake, for example, to calculate the percentage of a population that meets recommended dietary intake guidelines. We found that FFQs validated against recovery biomarkers underestimated the level of energy intake on average by 20 percent and the level of protein intake by 11 percent; thus, they are not suitable to assess levels accurately. However, FFQs are able to distinguish between subpopulations, as indicated by the analyses of the gender differences in energy intake. This difference was very similar to the gender difference found in a review that used doubly labeled water to estimate mean energy expenditure (75). These results showed that the average difference in energy intake was much smaller when standard portions were used. This is an argument to use at least sex-specific standard portions.

Our review shows that the number of items on the FFQ should not be reduced just because of the length of the food list; doing so might reduce the validity of the FFQ. In addition, portion-size questions do not improve FFQs for all nutrients. We should pay attention to this factor in the development process or by improving methods to estimate portion size. In addition, our review shows that FFQs are able to distinguish between subpopulations, although the magnitude of these differences is underestimated.

ACKNOWLEDGMENTS

This study was supported by the Netherlands Organization for Health Research and Development (ZonMw, grant 91104005).

Conflict of interest: none declared.

REFERENCES

1. Subar AF. Developing dietary assessment tools. *J Am Diet Assoc* 2004;104:769–70.
2. Schatzkin A, Kipnis V, Carroll RJ, et al. A comparison of a food frequency questionnaire with a 24-hour recall for use in an epidemiological cohort study: results from the biomarker-based Observing Protein and Energy Nutrition (OPEN) study. *Int J Epidemiol* 2003;32:1054–62.

3. Bingham SA, Luben R, Welch A, et al. Are imprecise methods obscuring a relation between fat and breast cancer? *Lancet* 2003;362:212–14.
4. Wolk A, Bergstrom R, Adami HO, et al. Self-administered food frequency questionnaire: the effect of different designs on food and nutrient intake estimates. *Int J Epidemiol* 1994;23: 570–6.
5. Jain M, McLaughlin J. Validity of nutrient estimates by food frequency questionnaires based either on exact frequencies or categories. *Ann Epidemiol* 2000;10:354–60.
6. Kuskowska Wolk A, Holte S, Ohlander EM, et al. Effects of different designs and extension of a food frequency questionnaire on response rate, completeness of data and food frequency responses. *Int J Epidemiol* 1992;21:1144–50.
7. Willett W. *Nutritional epidemiology*. 2nd ed. New York, NY: Oxford University Press, 1998.
8. Willett WC, Sampson L, Stampfer MJ, et al. Reproducibility and validity of a semiquantitative food frequency questionnaire. *Am J Epidemiol* 1985;122:51–65.
9. Block G, Woods M, Potosky A, et al. Validation of a self-administered diet history questionnaire using multiple diet records. *J Clin Epidemiol* 1990;43:1327–35.
10. Riboli E. Nutrition and cancer: background and rationale of the European Prospective Investigation into Cancer and Nutrition (EPIC). *Ann Oncol* 1992;3:783–91.
11. Kleinbaum D, Kupper L, Muller K, et al. *Applied regression analysis and multivariable methods*. 3rd ed. Pacific Grove, CA: Brooks/Cole Publishing Company, 1998.
12. Field AP. Is the meta-analysis of correlation coefficients accurate when population correlations vary? *Psychol Methods* 2005;10:444–67.
13. Thompson SG, Sharp SJ. Explaining heterogeneity in meta-analysis: a comparison of methods. *Stat Med* 1999;18:2693–708.
14. Bingham SA. Urine nitrogen as a biomarker for the validation of dietary protein intake. *J Nutr* 2003;133(suppl 3):921s–4s.
15. Matthews DE. *Modern nutrition in health and disease*. Baltimore, MD: Williams & Wilkins Company, 1999.
16. Bohlscheid Thomas S, Hoting I, Boeing H, et al. Reproducibility and relative validity of food group intake in a food frequency questionnaire developed for the German part of the EPIC project. *European Prospective Investigation into Cancer and Nutrition*. *Int J Epidemiol* 1997;26(suppl 1):S59–70.
17. Boucher B, Cotterchio M, Kreiger N, et al. Validity and reliability of the Block98 food-frequency questionnaire in a sample of Canadian women. *Public Health Nutr* 2006;9:84–93.
18. Flagg EW, Coates RJ, Calle EE, et al. Validation of the American Cancer Society Cancer Prevention Study II Nutrition Survey Cohort Food Frequency Questionnaire. *Epidemiology* 2000;11:462–8.
19. Jain MG, Rohan TE, Soskolne CL, et al. Calibration of the dietary questionnaire for the Canadian Study of Diet, Lifestyle and Health cohort. *Public Health Nutr* 2003;6:79–86.
20. Johansson I, Hallmans G, Wikman A, et al. Validation and calibration of food-frequency questionnaire measurements in the Northern Sweden Health and Disease cohort. *Public Health Nutr* 2002;5:487–96.
21. Katsouyanni K, Rimm EB, Gnardellis C, et al. Reproducibility and relative validity of an extensive semi-quantitative food frequency questionnaire using dietary records and biochemical markers among Greek schoolteachers. *Int J Epidemiol* 1997; 26(suppl 1):S118–27.
22. Ocke MC, Bueno de Mesquita HB, Pols MA, et al. The Dutch EPIC food frequency questionnaire. II. Relative validity and reproducibility for nutrients. *Int J Epidemiol* 1997;26(suppl 1): S49–58.
23. Pisani P, Faggiano F, Krogh V, et al. Relative validity and reproducibility of a food frequency dietary questionnaire for use in the Italian EPIC centres. *Int J Epidemiol* 1997; 26(suppl 1):S152–60.
24. Subar AF, Thompson FE, Kipnis V, et al. Comparative validation of the Block, Willett, and National Cancer Institute food frequency questionnaires: the Eating at America's Table Study. *Am J Epidemiol* 2001;154:1089–99.
25. van Liere MJ, Lucas F, Clavel F, et al. Relative validity and reproducibility of a French dietary history questionnaire. *Int J Epidemiol* 1997;26(suppl 1):S128–36.
26. Andersen LF, Solvoll K, Johansson LR, et al. Evaluation of a food frequency questionnaire with weighed records, fatty acids, and alpha-tocopherol in adipose tissue and serum. *Am J Epidemiol* 1999;150:75–87.
27. Bingham SA, Gill C, Welch A, et al. Validation of dietary assessment methods in the UK arm of EPIC using weighed records, and 24-hour urinary nitrogen and potassium and serum vitamin C and carotenoids as biomarkers. *Int J Epidemiol* 1997;26(suppl 1):S137–51.
28. Block G, Hartman AM, Naughton D. A reduced dietary questionnaire: development and validation. *Epidemiology* 1990;1:58–64.
29. Block G, Thompson FE, Hartman AM, et al. Comparison of two dietary questionnaires validated against multiple dietary records collected during a 1-year period. *J Am Diet Assoc* 1992;92:686–93.
30. Brunner E, Stallone D, Juneja M, et al. Dietary assessment in Whitehall II: comparison of 7 d diet diary and food-frequency questionnaire and validity against biomarkers. *Br J Nutr* 2001;86:405–14.
31. Callmer E, Riboli E, Saracci R, et al. Dietary assessment methods evaluated in the Malmö food study. *J Intern Med* 1993;233:53–7.
32. Engle A, Lynn LL, Koury K, et al. Reproducibility and comparability of a computerized, self-administered food frequency questionnaire. *Nutr Cancer* 1990;13:281–92.
33. Fidanza F, Gentile MG, Porrini M. A self-administered semi-quantitative food-frequency questionnaire with optical reading and its concurrent validation. *Eur J Epidemiol* 1995;11:163–70.
34. Goldbohm RA, van den Brandt PA, Brants HA, et al. Validation of a dietary questionnaire used in a large-scale prospective cohort study on diet and cancer. *Eur J Clin Nutr* 1994;48: 253–65.
35. Hartwell DL, Henry CJ. Comparison of a self-administered quantitative food amount frequency questionnaire with 4-day estimated food records. *Int J Food Sci Nutr* 2001;52: 151–9.
36. Larkin FA, Metzner HL, Thompson FE, et al. Comparison of estimated nutrient intakes by food frequency and dietary records in adults. *J Am Diet Assoc* 1989;89:215–23.
37. Longnecker MP, Lissner L, Holden JM, et al. The reproducibility and validity of a self-administered semiquantitative food frequency questionnaire in subjects from South Dakota and Wyoming. *Epidemiology* 1993;4:356–65.
38. Mannisto S, Virtanen M, Mikkonen T, et al. Reproducibility and validity of a food frequency questionnaire in a case-control study on breast cancer. *J Clin Epidemiol* 1996;49:401–9.
39. Martin-Moreno JM, Boyle P, Gorgojo L, et al. Development and validation of a food frequency questionnaire in Spain. *Int J Epidemiol* 1993;22:512–19.
40. McKeown NM, Day NE, Welch AA, et al. Use of biological markers to validate self-reported dietary intake in a random sample of the European Prospective Investigation into Cancer

- United Kingdom Norfolk cohort. *Am J Clin Nutr* 2001;74:188–96.
41. Mungar RG, Folsom AR, Kushi LH, et al. Dietary assessment of older Iowa women with a food frequency questionnaire: nutrient intake, reproducibility, and comparison with 24-hour dietary recall interviews. *Am J Epidemiol* 1992;136:192–200.
 42. Patterson RE, Kristal AR, Tinker LF, et al. Measurement characteristics of the Women's Health Initiative food frequency questionnaire. *Ann Epidemiol* 1999;9:178–87.
 43. Pietinen P, Hartman AM, Haapa E, et al. Reproducibility and validity of dietary assessment instruments. I. A self-administered food use questionnaire with a portion size picture booklet. *Am J Epidemiol* 1988;128:655–66.
 44. Pietinen P, Hartman AM, Haapa E, et al. Reproducibility and validity of dietary assessment instruments. II. A qualitative food frequency questionnaire. *Am J Epidemiol* 1988;128:667–76.
 45. Riboli E, Elmstahl S, Saracci R, et al. The Malmo Food Study: validity of two dietary assessment methods for measuring nutrient intake. *Int J Epidemiol* 1997;26(suppl 1):S161–73.
 46. Rimm EB, Giovannucci EL, Stampfer MJ, et al. Reproducibility and validity of an expanded self-administered semi-quantitative food frequency questionnaire among male health professionals. *Am J Epidemiol* 1992;135:114–26.
 47. Schroder H, Covas MI, Marrugat J, et al. Use of a three-day estimated food record, a 72-hour recall and a food-frequency questionnaire for dietary assessment in a Mediterranean Spanish population. *Clin Nutr* 2001;20:429–37.
 48. Tjonneland A, Overvad K, Haraldsdottir J, et al. Validation of a semiquantitative food frequency questionnaire developed in Denmark. *Int J Epidemiol* 1991;20:906–12.
 49. Willett WC, Sampson L, Browne ML, et al. The use of a self-administered questionnaire to assess diet four years in the past. *Am J Epidemiol* 1988;127:188–99.
 50. Feunekes GI, Van Staveren WA, De Vries JH, et al. Relative and biomarker-based validity of a food-frequency questionnaire estimating intake of fats and cholesterol. *Am J Clin Nutr* 1993;58:489–96.
 51. Kroke A, Klipstein Grobusch K, Voss S, et al. Validation of a self-administered food-frequency questionnaire administered in the European Prospective Investigation into Cancer and Nutrition (EPIC) Study: comparison of energy, protein, and macronutrient intakes estimated with the doubly labeled water, urinary nitrogen, and repeated 24-h dietary recall methods. *Am J Clin Nutr* 1999;70:439–47.
 52. Subar AF, Kipnis V, Troiano RP, et al. Using intake biomarkers to evaluate the extent of dietary misreporting in a large sample of adults: the OPEN study. *Am J Epidemiol* 2003;158:1–13.
 53. Andersen LF, Tomten H, Haggarty P, et al. Validation of energy intake estimated from a food frequency questionnaire: a doubly labelled water study. *Eur J Clin Nutr* 2003;57:279–84.
 54. Block G, Hartman AM, Dresser CM, et al. A data-based approach to diet questionnaire design and testing. *Am J Epidemiol* 1986;124:453–69.
 55. Bingham SA, Day NE. Using biochemical markers to assess the validity of prospective dietary assessment methods and the effect of energy adjustment. *Am J Clin Nutr* 1997;65(suppl):1130S–7S.
 56. Bohlscheid-Thomas S, Hoting I, Boeing H, et al. Reproducibility and relative validity of energy and macronutrient intake of a food frequency questionnaire developed for the German part of the EPIC project. European Prospective Investigation into Cancer and Nutrition. *Int J Epidemiol* 1997;26(suppl 1):S71–81.
 57. Kaaks R, Riboli E. Validation and calibration of dietary intake measurements in the EPIC project: methodological considerations. European Prospective Investigation into Cancer and Nutrition. *Int J Epidemiol* 1997;26(suppl 1):S15–25.
 58. Lietz G, Barton KL, Longbottom PJ, et al. Can the EPIC food-frequency questionnaire be used in adolescent populations? *Public Health Nutr* 2002;5:783–9.
 59. Bartali B, Turrini A, Salvini S, et al. Dietary intake estimated using different methods in two Italian older populations. *Arch Gerontol Geriatr* 2004;38:51–60.
 60. EPIC Group of Spain. Relative validity and reproducibility of a diet history questionnaire in Spain. II. Nutrients. European Prospective Investigation into Cancer and Nutrition. *Int J Epidemiol* 1997;26(suppl 1):S100–9.
 61. Riboli E, Toniolo P, Kaaks R, et al. Reproducibility of a food frequency questionnaire used in the New York University Women's Health Study: effect of self-selection by study subjects. *Eur J Clin Nutr* 1997;51:437–42.
 62. Willett W, Stampfer MJ. Total energy intake: implications for epidemiologic analyses. *Am J Epidemiol* 1986;124:17–27.
 63. Kipnis V, Subar AF, Midthune D, et al. Structure of dietary measurement error: results of the OPEN biomarker study. *Am J Epidemiol* 2003;158:14–21.
 64. Subar AF, Heimendinger J, Patterson BH, et al. Fruit and vegetable intake in the United States: the baseline survey of the Five A Day for Better Health Program. *Am J Health Promot* 1995;9:352–60.
 65. Feunekes GI, van't Veer P, van Staveren WA, et al. Alcohol intake assessment: the sober facts. *Am J Epidemiol* 1999;150:105–12.
 66. Noethlings U, Hoffmann K, Bergmann MM, et al. Portion size adds limited information on variance in food intake of participants in the EPIC-Potsdam study. *J Nutr* 2003;133:510–15.
 67. Burley V, Cade J, Margetts B, et al. Consensus document on the development, validation and utilisation of food frequency questionnaires. Nuffield Institute for Health, University of Leeds and Institute of Human Nutrition, University of Southampton, United Kingdom, 2000:63.
 68. Cade J, Thompson R, Burley V, et al. Development, validation and utilisation of food-frequency questionnaires—a review. *Public Health Nutr* 2002;5:567–87.
 69. Tjonneland A, Haraldsdottir J, Overvad K, et al. Influence of individually estimated portion size data on the validity of a semiquantitative food frequency questionnaire. *Int J Epidemiol* 1992;21:770–7.
 70. Young LR, Nestle M. Expanding portion sizes in the US marketplace: implications for nutrition counseling. *J Am Diet Assoc* 2003;103:231–4.
 71. Young LR, Nestle MS. Portion sizes in dietary assessment: issues and policy implications. *Nutr Rev* 1995;53:149–58.
 72. Wise A, Birrell NM. Design and analysis of food frequency questionnaires—review and novel method. *Int J Food Sci Nutr* 2002;53:273–9.
 73. Wirfalt AK, Jeffery RW, Elmer PJ. Comparison of food frequency questionnaires: the reduced Block and Willett questionnaires differ in ranking on nutrient intakes. *Am J Epidemiol* 1998;148:1148–56.
 74. Tyllavski FA, Sharp GB. Misclassification of nutrient and energy intake from use of close-ended questions in epidemiologic research. *Am J Epidemiol* 1995;142:342–52.
 75. Black AE, Cole TJ. Within- and between-subject variation in energy expenditure measured by the doubly-labelled water technique: implications for validating reported dietary energy intake. *Eur J Clin Nutr* 2000;54:386–94.